

Exercice 1. L'évolution de la mortalité en France au cours du 20^e siècle

CORRIGÉ

Entre 1900 et 2000, le taux de mortalité, qui mesure la proportion de personnes qui décèdent chaque année au sein d'une population, a diminué de moitié en France : il était de 20 ‰ (20 décès pour 1 000 personnes) en 1900 alors qu'il n'était plus que de 9 ‰ en 2000. Cette évolution est la conséquence de la réduction des risques de décéder à chaque âge, qui s'est également traduite par une augmentation considérable de l'espérance de vie à la naissance (de 45 à 79 ans).

L'évolution de la mortalité au cours du 20^e siècle est aussi caractérisée par une variation considérable de la distribution des décès selon l'âge. C'est l'objet de cet exercice.

Q.1 Calculez, pour les années 1900 et 2000, l'âge moyen des personnes qui sont décédées.

Q.2 Calculez, pour chacune de ces années, les quartiles (Q₁, Q₂ et Q₃). Explicitez le sens de chacun de ces indicateurs à l'aide d'une comparaison des résultats obtenus pour les années 1900 et 2000.

Q.3 Calculez, pour chacune de ces années, l'écart-type. Là encore, explicitez de manière concrète le sens de cet indicateur.

Q.4 À partir de ces caractéristiques de tendance centrale et de dispersion, commentez l'évolution de la distribution des décès selon l'âge au cours du 20^e siècle. Expliquez en particulier le lien entre l'évolution de la distribution des décès selon l'âge et la variation de l'âge moyen des personnes qui décèdent.

////////////////////////////////////

Corrigé

Q.1 & Q.3. Pour traiter cet exercice de manière efficace et logique, on pouvait commencer par calculer l'âge moyen des personnes décédées en 1900 et 2000 puis l'écart-type, puisque cette caractéristique de dispersion donne une mesure approchée de l'écart moyen des valeurs de la variable autour de la moyenne.

L'âge moyen est la moyenne des centres de classes d'âges pondérés par le poids des différentes catégories d'âges. Il s'écrit dans le cas présent :

$$\bar{x} = \sum_{i=5}^{95} (f_i \times x_i)$$

L'écart-type correspond à la racine carrée de la moyenne des carrés des écarts à la moyenne. On peut formuler cet indicateur de deux façons différentes :

$$\sigma_x = \sqrt{\sum_{i=5}^{95} [f_i \times (x_i - \bar{x})^2]}$$

ou

$$\sigma_x = \sqrt{\sum_{i=5}^{95} (f_i \times x_i^2) - \bar{x}^2}$$

Nous allons utiliser la seconde formulation de l'écart-type pour les calculs.

a) Tableau récapitulatif pour les décès en 1900

Groupe d'âges	Centre de classe x_i	Proportion f_i	$f_i \times x_i$	$f_i \times x_i^2$
0-9 ans	5	23,8%	1,2	6,0
10-19 ans	15	3,4%	0,5	7,5
20-29 ans	25	5,7%	1,4	35,0
30-39 ans	35	6,1%	2,1	73,5
40-49 ans	45	7,0%	3,2	144,0
50-59 ans	55	9,7%	5,3	291,5
60-69 ans	65	15,0%	9,8	637,0
70-79 ans	75	18,9%	14,2	1065,0
80-89 ans	85	9,6%	8,2	697,0
90 ans ou +	95	0,8%	0,8	76,0
Total		100,0%	46,7	3032,5

$$\bar{x} = (0,238 \times 5) + (0,034 \times 15) + \dots + (0,096 \times 85) + (0,008 \times 95) = 1,2 + 0,5 + \dots + 8,2 + 0,8 = 46,7$$

L'âge moyen des personnes décédées en 1900 est de 46,7 ans.

$$\sigma_x = \sqrt{[(0,238 \times 5^2) + (0,034 \times 15^2) + \dots + (0,096 \times 85^2) + (0,008 \times 95^2)] - 46,7^2}$$

$$\sigma_x = \sqrt{[(0,238 \times 5) \times 5 + ((0,034 \times 15) \times 15) + \dots + ((0,096 \times 85) \times 85) + ((0,008 \times 95) \times 95)] - 46,7^2}$$

$$\sigma_x = \sqrt{[(1,2 \times 5) + (0,5 \times 15) + \dots + (8,2 \times 85) + (0,8 \times 95)] - 46,7^2}$$

$$\sigma_x = \sqrt{[6,0 + 7,5 + \dots + 697,0 + 76,0] - 46,7^2}$$

$$\sigma_x = \sqrt{3032,5 - 46,7^2} = \sqrt{3032,5 - 2180,9} = \sqrt{851,6}$$

$$\sigma_x = 29,2$$

L'écart-type autour de la moyenne est de 29,2 ans.

b) Tableau récapitulatif pour les décès en 2000

Pour calculer l'âge moyen des personnes décédées en 2000 ainsi que la dispersion moyenne autour de cet âge moyen (écart-type), il est nécessaire au préalable de déterminer la répartition des décès en 2000 selon l'âge. Il faut donc « décumuler » la série des fréquences relatives cumulées croissantes présentées dans le tableau 2. Par exemple, si 1,4 % des personnes décédées en 2000 ont moins de 20

ans sachant que 0,9 % ont moins de 10 ans, on peut en déduire que la part des personnes âgées de 10-19 ans parmi l'ensemble des personnes décédées en 2000 est de 0,5 % :

$$f_{10-19} = f_{<20\text{ans}} - f_{<10\text{ans}}$$

$$f_{10-19} = 1,4\% - 0,9\% = 0,5\%$$

De manière générale, compte tenu des données disponibles dans le tableau, la part de chaque catégorie d'âges est obtenue par la différence suivante :

$$f_{i,i+9} = f_{<i+10\text{ans}} - f_{<i\text{ans}}$$

Groupe d'âges	Centre de classe x_i	Proportion f_i	$f_i \times x_i$	$f_i \times x_i^2$
0-9 ans	5	0,9%	0,0	0,0
10-19 ans	15	0,5%	0,1	1,5
20-29 ans	25	1,2%	0,3	7,5
30-39 ans	35	2,1%	0,7	24,5
40-49 ans	45	4,6%	2,1	94,5
50-59 ans	55	6,6%	3,6	198,0
60-69 ans	65	13,5%	8,8	572,0
70-79 ans	75	22,7%	17,0	1275,0
80-89 ans	85	31,8%	27,0	2295,0
90 ans ou +	95	16,1%	15,3	1453,5
Total		100,0%	74,9	5921,5

$$\bar{x} = (0,009 \times 5) + (0,005 \times 15) + \dots + (0,318 \times 85) + (0,161 \times 95) = 0,0 + 0,1 + \dots + 27,0 + 15,3 = 74,9$$

L'âge moyen des personnes décédées en 2000 est de 74,9 ans.

$$\sigma_x = \sqrt{[(0,009 \times 5^2) + (0,005 \times 15^2) + \dots + (0,318 \times 85^2) + (0,161 \times 95^2)] - 74,9^2}$$

$$\sigma_x = \sqrt{[((0,009 \times 5) \times 5) + ((0,005 \times 15) \times 15) + \dots + ((0,318 \times 85) \times 85) + ((0,161 \times 95) \times 95)] - 74,9^2}$$

$$\sigma_x = \sqrt{[(0,0 \times 5) + (0,1 \times 15) + \dots + (27,0 \times 85) + (15,3 \times 95)] - 74,9^2}$$

$$\sigma_x = \sqrt{[0,0 + 1,5 + \dots + 2295,0 + 1453,5] - 74,9^2}$$

$$\sigma_x = \sqrt{5921,5 - 74,9^2} = \sqrt{5921,5 - 5610,0} = \sqrt{311,5}$$

$$\sigma_x = 17,6$$

L'écart-type autour de la moyenne est de 17,6 ans.

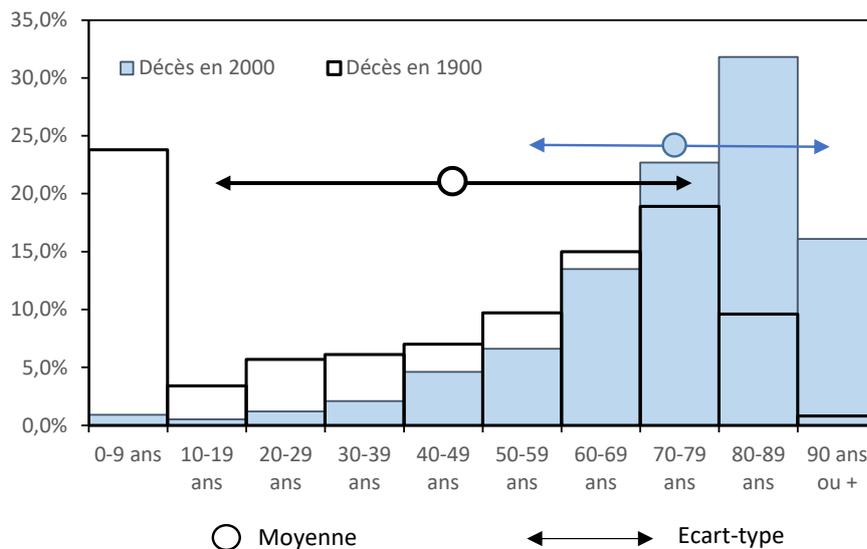
c) Comparaison

En 1900, les personnes décédées avaient en moyenne 46,7 ans, soit près de 30 ans de moins que celles décédées en 2000 (74,9 ans). Comme la limite d'âge au décès n'a guère évolué (autour de 105 ans), le recul de l'âge moyen des personnes qui décèdent s'accompagne nécessairement d'une plus forte concentration des décès autour de l'âge moyen : c'est ce que souligne l'évolution de l'écart-type, qui passe de 29,2 ans en 1900 à 17,6 ans en 2000.

Le graphique 1 montre clairement la concentration des décès aux âges élevés en 2000. L'âge moyen est très proche de la classe d'âge modale. Sur le plan statistique, on parle de distribution « étalée sur la gauche ». Dans ce cas, la moyenne est inférieure à la médiane (79,1 ans ; cf. infra le calcul de cet indicateur) qui est elle-même inférieure au mode (compris entre 80 et 90 ans).

Au contraire, la distribution des décès est beaucoup plus étalée en 1900. Cette distribution présente même deux classes d'âges modales très éloignées l'une de l'autre : celle des 0-9 ans et celle des 70-79 ans.

Graphique 1 : Répartition selon l'âge des décès dénombrés en 1900 et 2000



Q2. On peut également caractériser l'évolution de la mortalité à partir des quartiles.

Rappel :

- Le premier quartile (Q_1) partage la série statistique (ici les personnes décédées) en deux effectifs inégaux. La valeur de la variable (ici l'âge) prise par $\frac{1}{4}$ de l'effectif est inférieure au premier quartile ; le reste de l'effectif (soit les $\frac{3}{4}$) prend donc une valeur de la variable supérieure au premier quartile.
- La médiane (ou deuxième quartile : Me ou Q_2) partage la série statistique en deux effectifs égaux. La valeur de la variable prise par la moitié de l'effectif est inférieure à la médiane ; l'autre moitié de l'effectif prend donc une valeur de la variable supérieure à la médiane.
- Le troisième quartile (Q_3) partage la série statistique en deux effectifs inégaux. La valeur de la variable prise par les $\frac{3}{4}$ de l'effectif est inférieure au troisième quartile ; le $\frac{1}{4}$ de l'effectif restant prend donc une valeur de la variable supérieure au troisième quartile.

- Calcul détaillé des quartiles pour les décès en 1900

Pour déterminer les quartiles caractérisant la distribution des décès selon l'âge en 1900, il faut au préalable calculer les fréquences relatives cumulées. Par exemple, la proportion de personnes décédées avant l'âge de 20 ans correspond à la somme de la part de personnes décédées à 0-9 ans et de celle des personnes décédées à 10-19 ans :

$$f_{<20ans} = f_{0-9} + f_{10-19} = 23,8\% + 3,4\% = 27,2\%$$

27,2 % des personnes décédées en 1900 avaient moins de 20 ans.

D'une manière générale :

$$f_{<i \text{ ans}} = \sum_{x=0}^{i-1} f_x$$

Tableau récapitulatif

Groupe d'âges	Proportion f_i	f_i cumulée	
0-9 ans	23,8%	23,8%	moins de 10 ans
10-19 ans	3,4%	27,2%	moins de 20 ans
20-29 ans	5,7%	32,9%	moins de 30 ans
30-39 ans	6,1%	39,0%	moins de 40 ans
40-49 ans	7,0%	46,0%	moins de 50 ans
50-59 ans	9,7%	55,7%	moins de 60 ans
60-69 ans	15,0%	70,7%	moins de 70 ans
70-79 ans	18,9%	89,6%	moins de 80 ans
80-89 ans	9,6%	99,2%	moins de 90 ans
90 ans ou +	0,8%		
Total	100,0%		

– Calcul de Q_1 : Le premier quartile est compris entre 10 et 20 ans. En effet, 23,8 % des personnes décédées en 1900 avaient moins de 10 ans tandis que 27,2 % avaient moins de 20 ans.

$$Q_1 = 10 + \frac{(25\% - 23,8\%)}{(27,2\% - 23,8\%)} \times (20 - 10) = 13,5 \text{ ans}$$

Un quart des personnes décédées en 1900 avait moins de 13,5 ans.

– Calcul de Q_2 (ou Me) : La médiane est comprise entre 50 et 60 ans. En effet, 46,0 % des personnes décédées en 1900 avaient moins de 50 ans tandis que 55,7 % avaient moins de 60 ans.

$$Q_2 = 50 + \frac{(50\% - 46,0\%)}{(55,7\% - 46,0\%)} \times (60 - 50) = 54,1 \text{ ans}$$

La moitié des personnes décédées en 1900 avait moins de 54,1 ans.

– Calcul de Q_3 : Le troisième quartile est compris entre 70 et 80 ans. En effet, 70,7 % des personnes décédées en 1900 avaient moins de 70 ans tandis que 89,6 % avaient moins de 80 ans.

$$Q_3 = 70 + \frac{(75\% - 70,7\%)}{(89,6\% - 70,7\%)} \times (80 - 70) = 72,3 \text{ ans}$$

Les trois quarts des personnes décédées en 1900 avaient moins de 72,3 ans. On peut aussi dire qu'un quart des personnes décédées en 1900 avait plus de 72,3 ans.

- Calcul des quartiles pour les décès en 2000

Le tableau 2 indique déjà les fréquences relatives cumulées. On adopte la même démarche que pour les décès en 1900.

– $Q_1 = 60 + \frac{(25\% - 15,9\%)}{(29,4\% - 15,9\%)} \times (70 - 60) = 66,7 \text{ ans}$

$$- Q_2 = 70 + \frac{(50\% - 29,4\%)}{(52,1\% - 29,4\%)} \times (80 - 70) = 79,1 \text{ ans}$$

$$- Q_3 = 80 + \frac{(75\% - 52,1\%)}{(83,9\% - 52,1\%)} \times (90 - 80) = 87,2 \text{ ans}$$

Q4. Commentaires

La distribution des décès selon l'âge a radicalement changé en un siècle. En 1900, les décès les plus nombreux étaient ceux d'enfants âgés de moins de 10 ans : ils représentaient à eux seuls près de un décès sur quatre (23,8 %). En 2000, ils ne représentent plus que 1 % des décès. Aujourd'hui, 7 décès sur 10 concernent des personnes âgées de 70 ans ou plus ; près de un décès sur trois (31,8 %) correspond à une personne âgée de 80-89 ans. L'évolution de la mortalité est donc caractérisée par une très grande concentration des décès aux âges élevés : en 2000, la moitié des personnes qui sont décédées étaient âgées de 66,7 à 87,2 ans, soit un intervalle de 20,5 ans⁽¹⁾, tandis qu'en 1900, la moitié des décès s'étalait sur une plage trois fois plus importante (entre 13,5 et 72,3 ans, soit un intervalle de 58,7 ans). Ce phénomène, lié notamment à la réduction considérable des risques de mortalité aux très jeunes âges puis à la réduction des risques de mortalité au-delà de 60 ans, s'est concrétisé par une augmentation importante de l'âge moyen des personnes décédées, qui est passé en un siècle de 46,7 ans à 74,9 ans.

(1) Il s'agit de l'intervalle interquartile, soit la différence entre le troisième et le premier quartile.

Remarque sur l'utilisation des caractéristiques de tendance centrale et de dispersion dans les commentaires : il faut privilégier les indicateurs les plus concrets. On aurait pu ici utiliser l'écart-type pour montrer la réduction de la dispersion des décès selon l'âge. Toutefois, l'intervalle interquartile me semble plus simple à comprendre, c'est la raison pour laquelle j'ai préféré l'utiliser. Dans le cas présent, il est également intéressant d'utiliser le mode (classe d'âges modale) de chacune des deux distributions car leur comparaison souligne de manière spectaculaire la transformation radicale de la distribution des décès en un siècle.

Données de l'exercice

Tableau 1 : Âge des personnes décédées en 1900

Groupe d'âges	Proportion
0-9 ans	23,8%
10-19 ans	3,4%
20-29 ans	5,7%
30-39 ans	6,1%
40-49 ans	7,0%
50-59 ans	9,7%
60-69 ans	15,0%
70-79 ans	18,9%
80-89 ans	9,6%
90 ans ou +	0,8%
Total	100,0%

Source : INSEE

Note de lecture : l'âge de 23,8 % des personnes décédées en 1900 était compris entre 0 et 10 ans exacts.

Tableau 2 : Cumul selon l'âge des personnes décédées en 2000

Âge	Proportion
moins de 10 ans	0,9%
moins de 20 ans	1,4%
moins de 30 ans	2,6%
moins de 40 ans	4,7%
moins de 50 ans	9,3%
moins de 60 ans	15,9%
moins de 70 ans	29,4%
moins de 80 ans	52,1%
moins de 90 ans	83,9%

Source : INSEE

Note de lecture : en 2000, 0,9 % des décès correspondent à des personnes âgées de moins de 10 ans exacts.